

González, Xulia; Miles-Touya, Daniel

Article

Price dispersion, chain heterogeneity, and search in online grocery markets

SERIEs - Journal of the Spanish Economic Association

Provided in Cooperation with:

Spanish Economic Association

Suggested Citation: González, Xulia; Miles-Touya, Daniel (2018) : Price dispersion, chain heterogeneity, and search in online grocery markets, SERIEs - Journal of the Spanish Economic Association, ISSN 1869-4195, Springer, Heidelberg, Vol. 9, Iss. 1, pp. 115-139, <https://doi.org/10.1007/s13209-018-0172-6>

This Version is available at:

<https://hdl.handle.net/10419/195270>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/>

Price dispersion, chain heterogeneity, and search in online grocery markets

Xulia González¹  · Daniel Miles-Touya¹

Received: 1 December 2016 / Accepted: 6 January 2018 / Published online: 25 January 2018
© The Author(s) 2018. This article is an open access publication

Abstract This paper identifies patterns of cross-sectional and temporal price dispersion—in the Spanish online grocery retail market—and evaluates the extent to which search costs and chain heterogeneity explain such dispersion. We build a data set comprising 836,074 prices for the most popular grocery products sold online by Spanish national chains at different locations. Our results show that price dispersion is still present (albeit to a lesser extent) even after controlling for chain heterogeneity and that it persists over time. We structurally estimate search costs distributions for two different baskets of goods using price data while accounting for vertical product differentiation. The analysis suggests that the extent of search is low. According to our estimates more than two thirds of consumers do *not* compare prices among supermarkets. We also find that more frequently purchased products have lower search cost and also lower price–cost margins.

Keywords Price dispersion · Retail food market · Pricing strategies · Search costs

We thank J. L. Moraga, M. J. Moral, C. Pazó and two anonymous reviewers for their valuable comments and suggestions as well as Claudia Martínez-Boubeta for excellent research assistance. The paper has also benefited presentations at Munich (EARIE 43rd annual conference) and Alicante (XXX Jornadas de Economía Industrial conference). We thank Marta Esteve and Diego Kuperman from Soysuper for providing data and technical assistance. All remaining errors are our own responsibility. Financial support from Ministerio de Economía y Competitividad and FEDER (ECO2015-69334-R) is gratefully acknowledged.

✉ Xulia González
xgzlez@uvigo.es

Daniel Miles-Touya
dmiles@uvigo.es

¹ Facultad Economía, Campus Lagoas Marcosende, 36210 Vigo, Spain

JEL Classification D4 · D83 · L11**1 Introduction**

Price dispersion arises when stores in the same market set different prices for the same homogenous good (Hopkins 2008). Different sources of heterogeneity among sellers as well as information frictions explain much of the price dispersion observed in many markets (Lach 2002; Baye et al. 2006). First, store differentiation (e.g., discount retailers versus well-established national chains) may lead to some stores persistently selling their products at a lower price level than others in the same market. Second, when some consumers are perfectly informed about market prices while others are not, price dispersion could arise as a mixed-strategy equilibrium outcome (Stigler 1961; Varian 1980).¹ In this case, stores' relative positions in the price distribution change over time—which makes it more difficult for consumers to identify the lowest-price store (Chandra and Tappata 2011). These two accounts are not mutually exclusive and may, when taken together, explain observed price dispersion. In fact, price dispersion remains after store differentiation is controlled for (Lach 2002; Baye et al. 2006).

Price dispersion is a common feature of the grocery retail sector (Zhao 2006). Supermarket chains typically differ in the services offered (i.e., in terms of product assortment, shipping services, return policies, Web design, etc.), and the strategy of unannounced, short-term reductions in the prices of certain products (i.e., sales) is frequently employed. Thus chain differentiation and search frictions might explain price dispersion in grocery markets.

In the case of *online* shopping, store differentiation appears to be less relevant and consumer search costs look lower than the costs of a “physical” search (Janssen et al. 2007); hence one might expect a reduction of price dispersion in the online context.² Nonetheless, empirical evidence suggests that price dispersion remains even when consumers can easily access price information on homogeneous products from retailers' Web pages or price comparison websites (Brynjolfsson and Smith 2000; Baye et al. 2004a, b; Ellison and Ellison 2009).

The aim of this empirical paper is to shed some light on price dispersion and search behavior in the Spanish online grocery sector. In Spain, online grocery sales represent a low yet growing share of the total grocery market. Most brick-and-mortar chains have already committed to this *e-grocery* segment, and most supermarket chains are increasing the number of services offered online. At the same time, the last few years have seen the deployment in Spain of several online price comparison applications that have significantly increased the price information available to consumers.³ The

¹ Subsequent theoretical studies include Rosenthal (1980), Burdett and Judd (1983), Carlson and McAfee (1983), Stahl (1989), Dana (1994), Baye and Morgan (2001), Janssen and Moraga-González (2004), and Janssen et al. (2005).

² Consumer access to prices—through supermarket websites or price comparison sites—may reduce search costs both for online and for offline retail consumers, since customers can easily check prices before or even while they shop. Baye et al. (2013) examine the evolution of product search since the pre-Internet era.

³ E.g. *carritus.com*, *tudespensa.com*, *Soysuper.com*, and *komparing.com* (among others), with responsive web designs.

empirical analysis is based on price information obtained from the comparison site Soysuper.com. The data cover a period of 182 days—from 1 October 2013 through 31 March 2014—and include 836,074 price observations. However, this data source does not reveal either the quantities sold online or the market shares of the respective grocery chains.

We therefore follow the literature that estimates the search cost distribution when only prices are observed while accounting for chain heterogeneity. In particular, we follow the approach of Wildenbeest (2011) who, building on Hong and Shum (2006) and Moraga-González and Wildenbeest (2008), allows for the estimation of search costs using only price data while accounting for vertical product differentiation.

Competition among supermarket stores is highly localized—even for online shopping. Supermarket websites might differ among locations in terms of product availability and prices.⁴ Hence we consider four distinct markets that are geographically distant from each other: Madrid, Barcelona, Málaga, and Vigo. We follow the literature that deals with search costs in the grocery market and estimate those costs for two baskets of goods: one that includes frequently purchased products (beverages, breakfast and cereals, dairy products, pantry, and personal care and household) and another that includes only alcoholic drinks (beer, wine, and spirits), which we assume is purchased less frequently. Each basket includes identical products (at the bar-code level) in all stores. Yet to the extent that online grocery chains remain heterogeneous in terms of reputation and/or range and level of services, the baskets cannot be viewed as homogeneous products.

As a preview of our results, we observe that price dispersion is still present (albeit to a lesser extent) even after we control for chain heterogeneity, and it persists over time. We find that chain heterogeneity explains much of the observed price dispersion and also that supermarkets' relative positions in the price ranking do not remain constant over time. These observed patterns suggest that price dispersion may be due to chain differentiation and mixed strategies, which justifies estimating search costs based on a model of *utility* competition (as in Wildenbeest 2011).

The estimated extent of search is low: across markets, more than two thirds of consumers search just once despite the low cost of searching. Overall, these results are in line with similar studies of the retail food market in other countries (e.g., France, the United Kingdom). In addition, our findings also indicate that the products purchased more frequently tend to entail lower search costs and are likely also to have lower price–cost margins.

The rest of this paper proceeds as follows. In Sect. 2 we discuss several empirical studies that measure price dispersion and search cost. Section 3 describes our data, and Sect. 4 presents empirical evidence of price dispersion in Spanish grocery markets. In Sect. 5 we provide details on our model and estimation strategy; the estimation results follow in Sect. 6. We conclude with a summary of our findings in Sect. 7.

⁴ A shopper who logs on to a supermarket's website must check for whether delivery is available in her postal zone; since these zones are matched up to the nearest warehouse, prices (and also product availability) may well differ by location.

2 Empirical literature on price dispersion and search cost

There is a burgeoning empirical literature that seeks to explain the sources of price dispersion. On the one hand, there are studies measuring price dispersion for a single product category; examples include orange juice (Berck et al. 2008), gasoline (Barron et al. 2004; Chandra and Tappata 2011), books and CDs (Brynjolfsson and Smith 2000; Brynjolfsson et al. 2010), spare parts for cars (Delgado and Waterson 2003), computer and electronic products (Baye et al. 2004a, b; Ellison and Ellison 2009), and airline tickets (Bachis and Piga 2011; Orlov 2011).⁵

On the other hand, some studies compare price dispersion across different products with the goal of establishing empirical regularities that could help to identify the *sources* of price dispersion. For example, Lach (2002) compares goods of different price levels and finds that higher-priced products exhibit greater price dispersion. Sorensen (2000) compares prescription drugs purchased at different frequencies and reports that price dispersion is negatively correlated with the frequency of purchase.

The price data used in these studies are from different sources. However, recent years have seen a greater number of studies using Internet-sourced data—obtained either directly from retailer websites or indirectly from price comparison websites. See, for example, Clay et al. (2001), Brown and Goolsbee (2002), Baye et al. (2004a, b), Ellison and Ellison (2009), and Cavallo (2017, 2018).

The theoretical approach based on non-sequential consumer search has become increasingly popular in these empirical studies. Thus, for instance, Burdett and Judd (1983) propose a model under which price dispersion can be sustained in equilibrium provided that some consumers observe multiple prices while other consumers observe only one price; this asymmetric distribution of price information is attributed to search costs. Departing from this idea, Hong and Shum (2006) propose a model for estimating search cost distributions—by means of an empirical likelihood estimation procedure—when only price data are observed. Moraga-González and Wildenbeest (2008) modify Hong and Shum's approach by introducing a maximum likelihood estimator.

Wildenbeest (2011) introduces vertical product differentiation and search costs to explain price dispersion in the grocery retail industry. In that paper, search cost estimation is based on a basket of grocery items from the four leading UK retailers over a 12-week period 2008. In his model, each firm has its own price distribution because firms are heterogeneous in terms of overall quality; this setup accounts not only for price dispersion across firms but also for the observation that some firms have persistently higher average prices than others. Wildenbeest's findings suggest that nearly two thirds of the observed price variation is explained by supermarket heterogeneity. Hence, ignoring the vertical product differentiation component could lead to overestimation of search frictions. In addition, search intensity is low: 7 in 10 consumers visit only one supermarket.

Following Wildenbeest's (2011) approach, Richards et al. (2016) use online grocery price data from four large retailers in the United Kingdom to estimate search costs for consumers who engage in multi-product shopping. Their results confirm the finding of

⁵ For a review of the research on price dispersion, see Baye et al. (2006).

low search intensity; 4 out of 5 consumers search only once when shopping for products in multiple categories, and even larger share search only once when shopping for a single category of product.

Hortaçsu and Syverson (2004) analyze the mutual funds industry while assuming that consumers have identical tastes but different search costs. De los Santos et al. (2012) utilize a large data set on Web-browsing and purchasing behavior to test how well various search models capture actual consumer search behavior. Dubois and Perrone (2015) extend the model of Hortaçsu and Syverson to allow for heterogeneous consumer preferences; thus, products are differentiated both vertically and horizontally. Dubois and Perrone use also data derived from observed shopping behavior, since they examine all store visits made by households within a certain period of time. Seiler (2013) proposes a structural model, with imperfect information, that accounts for both inventory holdings and searching. This author uses a consumer-level panel data set (Kantar Worldpanel UK) on laundry detergent and reports that consumers are unaware of its price on 70% of their shopping trips that involve purchasing that product.

In short, results reported in the empirical literature on search costs suggest that consumer search behavior is limited.

3 Data description

Our data set consists of daily posted prices (in euros) for a total of 237 branded products in the categories of beverages, breakfast and cereals, dairy products, pantry, and personal care and household; we excluded fresh food and store brands to ensure comparability between supermarkets. The product prices were provided by the online price aggregator Soysuper.com and were selected based on that site's popularity index.⁶

The data set identifies products in great detail. Two products are considered to be different if they have different bar codes; for example, whole milk and low-fat milk of the same brand are viewed as different products, for which separate prices are recorded. We also distinguish between the same product being sold either in packs or individually. We identify each chain–product–store combination with a specific product code (SPC); thus, for example, a 500g box of Corn Flakes Kellogs stocked by Auchan and Carrefour in Madrid and Barcelona is represented by four distinct SPCs, each with its own time series of daily prices. Hence identical products have different SPCs depending on the store at which they are sold.

Prices were recorded daily, from 1 October 2013 through 31 March 2014 (182 days), at four locations across Spain: Madrid, Barcelona, Málaga, and Vigo.⁷ The total

⁶ This index is based on three factors: the frequency that a product is purchased in comparison with all products purchased online; the frequency with which a product is added to the basket as compared with all products added to baskets; and the number of supermarkets selling the focal product on their respective Web pages.

⁷ Madrid and Barcelona are home to (respectively) 3,165,235 and 1,602,386 inhabitants. Málaga and Vigo are located, respectively, in the south (Andalucía region) and northwest (Galicia region) of Spain and have 566,913 and 294,997 residents. Average annual income per capita among the four cities differs, but not by much: €35,090 in Barcelona, €36,636 in Madrid, €24,405 in Málaga, and €29,654 in Vigo.

number of price observations is 836,074. These prices were collected from the product postings on websites of the main Spanish supermarket chains (Auchan, Carrefour, El Corte Inglés, Eroski, and Mercadona)⁸ and of a regional chain (Condis) that operates only in Catalonia (see Table 8 in the “Appendix”).

The average Spanish household’s expenditures on food products amounted to €123.50 per month in 2014.⁹ According to the Report on Food Consumer Habits in Spain (2014) from the Spanish Agriculture and Food Ministry (MAPAMA 2014), consumers’ shopping habits depend on which products they want to buy. Fresh products are purchased more often, typically at specialized shops or small supermarkets, and consumers are more concerned about the quality than the price of such products. In contrast, the standardized products purchased during “primary shopping” (beverages, dairy, cereals, etc.) are often bought in large quantities and at larger stores, such as supermarkets or superstores.

We estimate search costs for two shopping baskets: the basic basket that includes 30 products included in all categories and the occasional one that includes only alcoholic beverages. Tables 9 and 10 list the items included in each basket as well as their prices. We assume that the basic basket are purchased with higher frequency than the occasional one.

4 Evidence of price dispersion

Varian (1980) was among the first to distinguish between spatial and temporal price dispersion. *Spatial* price dispersion implies that the firm’s place in the price distribution does not necessarily change over time. In our context, any permanent price differences could be explained by store differentiation as price dispersion should diminish over time if consumers can learn (from experience) that some firms consistently offer lower prices than others. *Temporal* price dispersion arises when the store’s position in a price-based ranking changes over time. In what follows we discuss both types of price dispersion in our data.

4.1 Spatial price dispersion

We measure price dispersion as the (log of) deviations in price from the daily mean (Lach 2002); that is, $g_{ijlt} = \log p_{ijlt} - \log \bar{p}_{ilt}$. Here p_{ijlt} denotes the price of item i in chain j at location l on day t , and, \bar{p}_{ilt} is the *mean* price across stores of item i at location l on day t , for $i = 1, \dots, 237$, $j = 1, \dots, 6$, $l = 1, \dots, 4$, and $t = 1, \dots, 182$.

Part of the dispersion observed in prices may be explained by product differentiation. Although we compare goods with the same physical characteristics, these goods are sold at different stores and also on different days. As a result, the products are not

⁸ Aggregate revenue of the five national chains accounted for 72.2% of the net sales of main food chains (CNC 2011).

⁹ See the Report on Food Consumer Habits in Spain (MAPAMA 2014) from the Spanish Agriculture and Food Ministry. Fresh products (fresh fruit, vegetables, meat, fish, eggs, and fresh bread) represent a 44.5% of those expenditures, which means that €69.5 were spent each month on non-fresh food products.

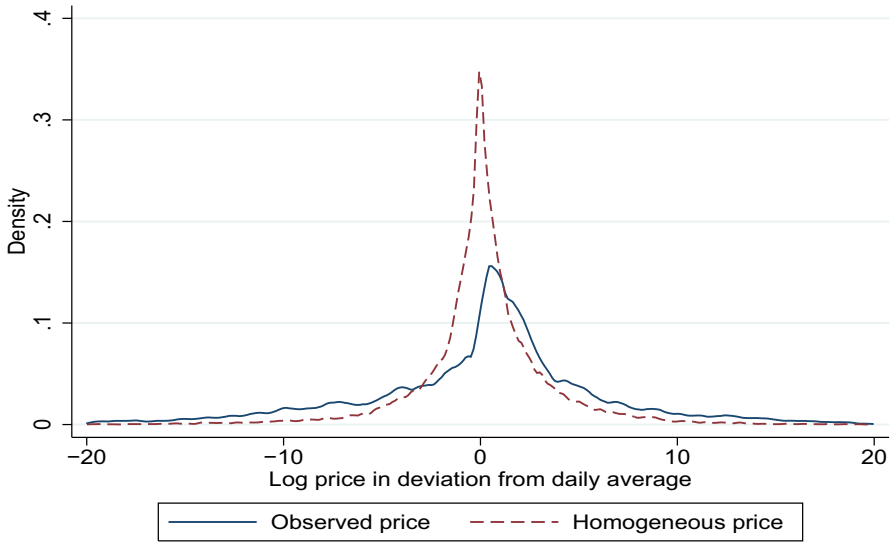


Fig. 1 Kernel density of spatial price dispersion utility (Barcelona)

homogenous. We therefore assume that identical products in different locations are *not* substitutes—since the distance between cities is great enough that we can view them as distinct markets.

Following Lach (2002), we remove from prices any heterogeneity due to the supermarket chain, location, or time. For this purpose we run item-by-item regressions of prices (measured as log deviations from the daily mean) on the fixed effects of supermarket chain and location, on the interaction between those effects, and on a time effect. We estimate the following equation for each item i , $g_{ijlt} = \mu + \alpha_j + \beta_l + \alpha_j \times \beta_l + \gamma_t + \varepsilon_{ijlt}$; here α_j is a supermarket-chain fixed effect that is common across products and locations, β_l is a location fixed effect, and γ_t is a time effect. The residuals of these regressions represent the price of a homogeneous product after controlling for time-invariant, store-specific effects and for the price fluctuations common to all stores. This method of deriving homogeneous prices is now standard in the literature (see e.g. Sorensen 2000; Lach 2002; Zhao 2006; Dubois and Perrone 2015).

Figure 1 plots the empirical density function for the spatial dispersion of observed prices (solid line) and of homogeneous prices pooled over products, stores, and days (dashed line).¹⁰

This figure reveals that prices exhibit considerable dispersion. Most observed prices fluctuate between -10% and $+10\%$ of the mean and, as expected, are relatively more dispersed than are the homogeneous prices. The empirical literature has documented that the magnitude of price dispersion varies as a function of product characteris-

¹⁰ Because our measure of price dispersion averages to zero each day, all variation in the kernel density is cross-sectional for each day and at each location.

Table 1 Moments of the price dispersion distribution (by category)

Categories	SD		Posted price		Homogeneous price	
	g_{ijlt}	ε_{ijlt}	p5	p95	p5	p95
Beverages	6.92	4.41	− 12.7	9.2	− 6.7	6.4
Breakfast and cereals	8.12	3.27	− 15.3	12.4	− 4.0	4.4
Coffee and cocoa	6.48	4.42	− 10.7	10.3	− 6.9	6.2
Dairy products	5.63	3.49	− 9.2	7.1	− 5.2	4.6
Pantry	9.54	4.44	− 17.6	12.8	− 6.0	6.2
Household and personal care	7.26	3.21	− 12.8	12.6	− 4.6	4.6

tics (price level, purchase frequency, etc.). Table 1 reports several measures of price dispersion by product categories.

The table's first two data columns give the standard deviations of the observed and homogeneous (residual) prices, and the last four columns report the 5th and 95th percentiles of each distribution. For all categories, spatial price dispersion drops significantly—in terms of both the standard deviation and the quantiles—when we control for observed product heterogeneity. Nevertheless, some dispersion persists even when prices are homogeneous. Results are somewhat heterogeneous with respect to product categories. The pantry category and the breakfast and cereals category exhibit the most price dispersion, since their standard deviations are the highest; prices for dairy products exhibit the least dispersion.

4.2 Temporal price dispersion

We observe temporal price dispersion when the identity of the store offering the lowest (and highest) price varies over time. If stores' relative positions in the price distribution remain constant, then dispersed prices could simply reflect store heterogeneity. However, if the ranking by prices varies over time then price dispersion might rather be the result of costly consumer search. In this latter case, firms change their prices so that buyers cannot identify the “cheapest” store overall.

Temporal price dispersion is measured by comparing how stores' prices are ranked over time. We define *rr* (rank reversal) as the number of changes in price-rank position for a given product i sold at supermarket j over τ days.¹¹ Thus, we write, $rr_i^j = \frac{1}{T} \sum_{t=\tau+1}^T I(r_{it}^j \neq r_{it-\tau}^j)$; here r_{it}^j denotes the position of the supermarket j in the price ranking of item i on day t in a given location, and $r_{it-\tau}^j$ denotes its position τ days earlier. Now we can define the *average ranking change* for supermarket j at a given location as the average of rank changes among all products: $RR^j = \frac{1}{N} \sum_{i=1}^N rr_i^j$. Although rank changes on two successive days may be small, the extent of change increase with the time interval.

¹¹ See Chandra and Tappata (2011) for a similar rank-reversal statistic.

Table 2 Average change in price ranking at four locations

	Barcelona			Madrid			Málaga			Vigo		
	<i>RR</i> ₁	<i>RR</i> ₁₀	<i>RR</i> ₃₀	<i>RR</i> ₁	<i>RR</i> ₁₀	<i>RR</i> ₃₀	<i>RR</i> ₁	<i>RR</i> ₁₀	<i>RR</i> ₃₀	<i>RR</i> ₁	<i>RR</i> ₁₀	<i>RR</i> ₃₀
Auchan	1.3	9.8	15.3	1.0	6.4	9.6	1.4	9.8	14.2	1.4	9.0	12.1
Carrefour	2.9	21.4	35.3	2.6	18.6	32.4	2.7	19.4	33.2	3.4	24.8	38.0
Condis	3.0	22.8	37.1									
El Corte Inglés	3.0	21.9	36.8	2.8	19.6	33.6	3.0	21.3	36.8	3.5	25.5	38.9
Eroski	3.3	23.1	38.5	2.8	19.9	33.9	3.1	21.5	36.3	3.5	25.8	39.4
Mercadona	2.8	21.1	35.1	2.6	18.3	32.0	2.9	20.9	34.2	3.5	25.4	39.6

Table 2 reports the average rank changes in prices between two observations with a time lag of 1, 10, and 30 days in Barcelona, Madrid, Málaga, and Vigo. A change in ranking position indicates a price change in at least one chain’s stores. However, a given supermarket can change a price without its rank changing and can also see its rank change without changing any prices—that is, provided some competitor does.

The table suggests that stores change their relative position in the price rankings as the time lag increases. The four first columns (one under each location) show that stores price rankings barely change between two consecutive days. Yet with a 10-day (resp. 30-day) time lag, the average percentage change approaches or exceeds 20% (resp. 35%). The exception to this pattern is Auchan, which is in the same position of the ranking most of the time. We shall see that their prices are always the lowest.¹²

Price rankings that fluctuate make it more difficult for consumers to find the best deal, and chances are nearly 4 in 10 that relative prices for a given product will have changed within the past month. It follows that consumers must update their information about prices rather frequently if they want to continue paying the least amount possible.

5 Using price data to estimate consumer search costs

The empirical evidence described so far suggests that part of the dispersion observed in prices may be explained by store differentiation, although our evidence is consistent also with imperfect information about prices. To the extent that supermarkets change their prices and thus their positions in the ranking, it is more difficult for consumers to learn about prices. In this section we estimate search costs as well as the proportion of consumers who compare prices when grocery shopping; for that purpose we use the available information about prices while accounting for supermarket chain heterogeneity.

We follow previous papers in this literature by considering the price of a basket of products rather than individual product prices, and we assemble two different baskets of

¹² Annual reports on supermarket prices published by the Spanish consumer organization (OCU) indicate that Auchan has been the lowest-price national chain in Spain: <https://www.ocu.org/consumo-familia/supermercados/informe/cadenas-mas-baratas> and https://elpais.com/economia/2017/09/25/actualidad/1506360054_741892.html (both accessed 29 September 2017).

goods. The *basic* basket includes those branded products—from various categories—most often included on consumer shopping lists and the *occasional* basket includes only alcoholic beverages. We estimate the search costs for the two baskets in four geographical markets.

5.1 The model

In estimating search costs, we use the nonsequential search model developed by Wildenbeest (2011)¹³. In this model firms compete directly in the utility space, which implies that the supermarket strategy space is reduced from two dimensions (quality and price) to a single utility dimension. This approach enables us to incorporate chain differentiation into the model.

Suppose there are N supermarkets offering a homogeneous good (a basket of groceries) to imperfectly informed consumers at a particular location. Supermarkets sell this “good” at a unit cost of r_j .

Consumers share a common utility function and have identical preferences regarding quality; however, their search costs differ (as in Hortaçsu and Syverson 2004). We can write

$$u_j = v_j - p_j \quad \text{for } j = 1, \dots, N,$$

where v_j is the consumer’s valuation of buying the good from supermarket j at a given location. This valuation has the additively separable structure $v(s_j) = x + s_j$; here x denotes the common consumers’ valuation of the homogeneous good independent of store quality, s_j is the supermarket’s level of services or quality (Wildenbeest 2011), and p_j is the basket’s price.

Because the consumers in our setup all have the same utility function, a supermarket determines its quality level s_j by maximizing the price–cost margin: $p_j - r_j = p(s_j) - r(s_j) = v(s_j) - r(s_j) - u$, for a given utility level u . By Euler’s theorem, the total cost of quality inputs exhausts quality-related output; that is, $r(s_j) = s_j$.¹⁴ As a consequence, the valuation cost markup does *not* depend on store quality: $v(s_j) - r(s_j) = x + s_j - r(s_j) = x$. We can therefore focus on symmetric mixed-strategy equilibria in utility levels, where the supermarket’s strategy is given by a common utility distribution function $L(u)$.

Consumers search nonsequentially. That is, a consumer takes supermarkets’ strategies as given and decides on the optimal number $k \geq 1$ of stores to visit, after which he buys from the store that gives him the highest utility.¹⁵ A consumer’s search cost c is assumed to be a random draw from a common atomless distribution $G(c)$ with support $(0, \infty)$ and positive density $g(c)$.

Consumer search behavior should be optimal in this sense: the net benefit of searching k times should be greater than that of searching either $k - 1$ or $k + 1$ times; and

¹³ See Moraga-González et al. (2017) for an analysis of the equilibrium’s existence and uniqueness.

¹⁴ Under perfect competition, supermarkets obtain quality input factors and the quality production function exhibits constant returns to scale.

¹⁵ A condition that partially characterizes a symmetric equilibrium in the utility space is that some consumers search only once (at no cost) while others search more than once (see Burdett and Judd 1983).

the expected utility from searching should exceed its expected cost ($k \cdot c$). The search cost of a consumer who is indifferent between searching k and $k + 1$ times,

$$c_k = Eu_{1:k+1} - Eu_{1:k} \text{ for } Eu_{1:k} = E[\max(u_1, \dots, u_k)],$$

is decreasing in k . The share q_k of consumers who sample k prices is then

$$q_k = \int_{c_k}^{c_{k-1}} g(c) dc = G(c_{k-1}) - G(c_k),$$

which implies:

$$\begin{aligned} q_1 &= G(c_0) - G(c_1) = 1 - G(c_1), \\ q_k &= G(c_{k-1}) - G(c_k) = 1 - \sum_{k=1}^{k-1} q_k - G(c_k) \text{ for } k = 2, \dots, N - 1, \\ q_N &= G(c_{N-1}). \end{aligned}$$

Here $G(c_0) = 1$, so every search cost is lower than c_0 ; and $G(c_N) = 0$, so c_N is the minimum search cost.

Supermarket j 's expected profit from offering utility level u_j is

$$\pi_j(u_j; L(u)) = (x - u_j) \sum_{k=1}^N \frac{k}{N} q_k L(u_j)^{k-1}$$

given expected consumer behavior q_k and the distribution function $L(u)$. Here $x - u_j = p_j - r_j$ is the price–cost margin of each supermarket that implicitly sets a price p . The summation captures the expected sales quantity, which depends on: (i) the proportion q_k of consumers searching k times; (ii) the likelihood k/N of consumers observing the utility of firm j ; and (iii) the probability $L(u_j)^{k-1}$ that, at utility level u_j , the firm offers the highest utility level of all the k firms searched.

In this setting, a price dispersion equilibrium is possible only when there exists a positive (though not certain) likelihood of a consumer observing only one price.¹⁶ The characterization of the equilibrium utility distribution in mixed strategies implies that the supermarket should be indifferent about what level of utility to set in the support of $L(u)$ (Burdett and Judd 1983). In particular, the supermarket should be indifferent between (a) offering a utility of zero by setting $\bar{p}_j = v_j$ and selling only to uninformed consumers (i.e., those who search just once) and (b) setting any other utility level in the support of $L(u)$, $u > 0 = \underline{u}$:

$$(x - u) \sum_{k=1}^N \frac{kq_k}{N} L(u)^{k-1} = x \frac{q_1}{N},$$

¹⁶ The intuition for this claim is that, if all consumers did compare the prices at different stores, then each firm would set a price equal to their unit cost.

where the right-hand side is the expected profit when offering zero utility. In this case, if firms offer the maximum utility ($u = \bar{u}$) then $L(\bar{u}) = 1$ (all utility levels are below the maximum); hence the maximum utility is given by

$$\bar{u} = x \frac{\sum_{k=2}^N kq_k}{\sum_{k=1}^N kq_k}.$$

Our aim is to estimate the points $\{q_k, c_k\}_{k=1}^n$ via maximum likelihood, as in Moraga-González and Wildenbeest (2008). The equilibrium distribution of utilities, $L(u)$, can only be implicitly defined, but the density function can be derived from the first-order conditions of expected profit maximization:

$$\frac{\partial \pi}{\partial u} = 0 \quad \text{and} \quad l(u) = \frac{\sum_{k=1}^N kq_k(L(u))^{k-1}}{(x - u) \sum_{k=2}^N k(k - 1)q_k(L(u))^{k-2}};$$

these conditions are then used to define the log-likelihood function $LL = \sum_{i=2}^T \log l(u_i)$. Here T is the total number of observations, the minimum utility is zero, and all utilities are arranged in ascending order: $\underline{u} = u_1 < u_2 < \dots < u_T = \bar{u}$. Thus we have $L(\bar{u}) = 1$ and $L(\underline{u}) = 0$.

We can use this characterization of optimal searching behavior to rewrite the search cost as follows:

$$\begin{aligned} c_k &= \int_{\bar{u}, \underline{u}} (k + 1)uL(u)^k l(u) du - \int_{\bar{u}, \underline{u}} kuL(u)^{k-1} l(u) du \\ &= \int_{\bar{u}, \underline{u}} u[(k + 1)L(u)^k - kL(u)^{k-1}]l(u) du \\ &= \int_{\bar{u}, \underline{u}} u[(k + 1)L(u) - k]L(u)^{k-1} l(u) du. \end{aligned}$$

Further simplification is possible if we put $y = L(u)$, so that $dy = l(u) du$ and $u = L^{-1}(y) = u(y)$; then

$$c_k = \int_0^1 u(y)[(k + 1)y - k]y^{k-1} dy.$$

If we now use the same change of variable in the equilibrium profit equation y yields

$$(x - u) \sum_{k=1}^N \frac{k\mu_k}{N} y^{k-1} = x \frac{\mu_1}{N},$$

from which the equality $u(y) = x - \frac{x\mu_1}{\sum_{k=1}^N k\mu_k y^{k-1}}$ follows¹⁷.

¹⁷ Moraga-Gonzalez et al. (2013) show that the model described here is nonparametrically identified if the number of searches does not tend to infinity.

5.2 Estimation strategy

The first step is to estimate utilities. Toward that end, we assume that consumers differ in their search costs but have the same preferences regarding chain characteristics.¹⁸ Thus consumers in a given location derive utility from buying the homogeneous basket at store j according to $u_j = v_j - p_j$, where v_j is the valuation of buying the basket at store j and p_j is that basket’s price. Utilities are then defined as prices adjusted by the heterogeneity between stores (services provided, quality, etc.). Consumers know their valuation of a good but do not know its price. Hence they must obtain information about basket prices at a number of supermarkets in accordance with their search costs.

We rewrite the preceding paragraph’s equation as $p_j = v_j - u_j$, which can be estimated via a fixed-effects regression on prices. Then $p_j = \alpha + \gamma_j + u_j$, where α is a constant, γ_j is a store fixed effect, and the (negative of) disturbance u_j represents utility.

After deriving the estimated utilities, we proceed as in Moraga-González and Wildenbeest (2008).¹⁹ The maximum likelihood (ML) problem is given by

$$\max_{\{q_k\}_{k=1}^{N-1}} \sum_{m=2}^M \log l(u_m; q_1, \dots, q_N).$$

The term M is the number of price data points at each location, and $L(u_j)$ solves

$$(x - u_j) \sum_{k=1}^N \frac{kq_k}{N} L(u_m)^{k-1} = x \frac{\mu_1}{N} \quad \text{for all } m = 2, 3, \dots, M - 1$$

and

$$x = \bar{u} \frac{\sum_{k=1}^N kq_k}{\sum_{k=2}^N kq_k};$$

and using the fact that $q_N = 1 - \sum_{k=1}^{N-1} q_k$. That is, the ML estimator yields estimates of the proportion q_k ($k = 1, \dots, N$) of consumers who are searching, and we can recover the search costs from those estimates.²⁰

¹⁸ This assumption implies, for example, that all consumers would prefer shopping at Eroski or El Corte Inglés than at Auchan if prices were identical across supermarket chains.

¹⁹ The estimates are derived while assuming that firms play a stationary repeated game of finite horizon, so the data in each period should reflect the equilibrium of the static game analyzed previously (cf. Moraga-González and Wildenbeest 2008). For the estimation we select price observations every 10 days to avoid serial correlation.

²⁰ The maximum likelihood estimator was programmed using MATLAB and the optimization procedure with its TOMLAB solver package.

Table 3 Price of the basic basket

Supermarket	Price			
	Mean	SD	Max.	Min.
Auchan	61.59	1.32	63.50	57.55
Carrefour	67.30	0.20	67.72	66.79
Condis	67.71	0.56	68.48	66.77
ElCorteInglés	68.09	0.57	69.32	65.76
Eroski	68.47	0.38	69.16	67.15
Mercadona	67.11	0.78	67.87	64.02
Average	66.57	2.58	69.32	57.55

6 Results

In order to estimate search costs we select two different baskets of products. First, a “basic” basket that includes the most frequently purchased grocery products. And the “occasional” basket, a basket containing only alcoholic beverages (wine, beer, and spirits).²¹ These products are assumed to be purchased less frequently and most of them are not included in the basic basket.

6.1 Price dispersion: basic basket

The basic basket of goods includes regularly purchased items (according to Soy-super.com’s popularity index) in five product categories (beverages, breakfast and cereals, dairy products, pantry, and household and personal care), for which we have a complete series of prices across all the chains at all locations. Because each product in this basket is branded, we can view the basket as a completely homogenous good. Supermarkets are assumed to be especially interested in the pricing of these highly popular products.

Our basic shopping basket (which excludes *fresh* groceries) contains 30 items, most of which are purchased as part of primary shopping. The *basket price* is the sum of all individual prices weighted by category—that is, according to the household monthly expenditures reported by the Spanish Household Budget Survey, 2013. Its total cost range from €57.55 (Auchan) to €69.32 (El Corte Inglés) and averages about €67 (see Table 3).²² The difference between the most and the least expensive basket is €11.80, but this difference varies across supermarkets. Auchan has the greatest intra-firm price dispersion, ranging between €58 and €64; the basket price at Carrefour exhibits very

²¹ Products included in the baskets must be sold in all supermarkets and in all locations during the entire study period (else comparisons would not be possible). We exclude not only products that are almost identical (e.g., we include low-fat milk but neither whole milk nor skim milk) but also different sizes of the same product (e.g., 2l vs. 500ml Coca-Cola). When confronted with similar products, we made a random selection to determine which should be included in the basket.

²² Table 9 (in the “Appendix”) reports the average price and the price dispersion of each product included in the basic basket.

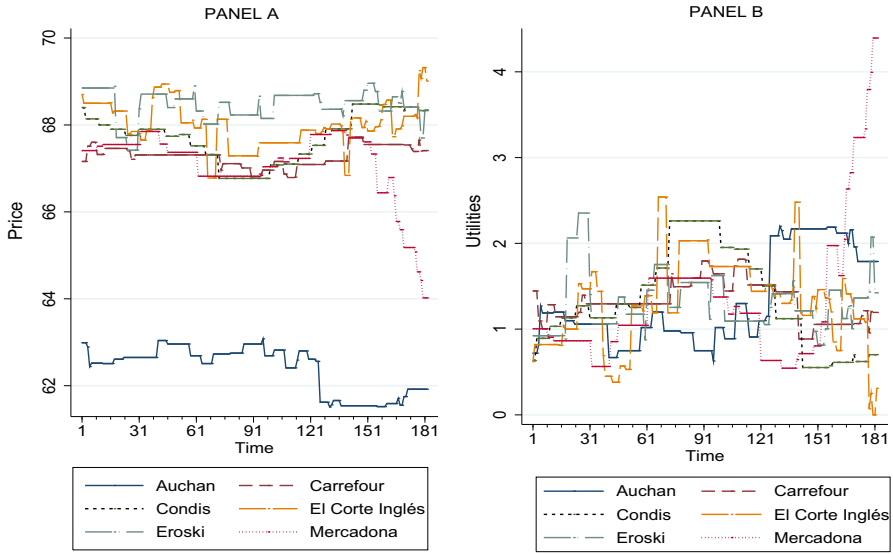


Fig. 2 Ranking over time, by price and estimated utility (Barcelona)

little variation (€66.80–€68.50). El Corte Inglés and Eroski have the highest prices. Price differences among stores belonging to the same chain are small in all chains except Auchan. Mercadona is the chain that exhibit lower intra-chain price dispersion, which suggests that most of its prices are centralized at the chain level.²³

If store characteristics do not change (at least in the short term) and if price dispersion reflects chains differentiation, then we should expect supermarkets to set prices in a manner that preserves their ranking position. Thus, high-quality supermarkets will nearly always set relatively high prices. Of course, if the position of stores in the price distribution remains constant then it is easier for consumers to learn about prices.

In Fig. 2, Panel A provides information on the daily price of the basic basket for the six supermarkets in Barcelona. Auchan is clearly the supermarket with the lowest basket price, yet its prices change frequently. The ranking of the other supermarkets varies, although El Corte Inglés usually has the highest prices and Mercadona the lowest—especially in the last month sampled, when the latter chain initiated a price-cutting campaign. Stores revise their pricing often but do not change prices synchronously; that is, the length of time between price changes differs across chains. In sum, Panel A suggest that some chains have consistently lower prices (and/or prices that change more frequently) than other chains. Panel B displays information on consumer utility, which is estimated as the negative of the residuals from a regression of prices on store fixed effects; the resulting value can be interpreted as the price of the homogeneous good after controlling for chain heterogeneity.²⁴ This panel shows

²³ See González (2017) for a discussion of pricing strategies at the chain level.

²⁴ We normalize minimum utility to zero adding a constant. Estimates of the search cost distribution do not change when a constant is added because the differences between utilities remain invariant.

Table 4 Periods spend by prices and utilities in each quartile (Barcelona)

Supermarket	Prices quartile				Utilities quartile			
	q1	q2	q3	q4	q1	q2	q3	q4
Auchan	100	0	0	0	32.4	32.4	3.8	31.3
Carrefour	4.9	76.9	18.1	0	6.6	26.9	52.7	13.7
Condis	14.8	24.2	34.1	26.98	26.9	26.4	15.9	30.8
El Corte Inglés	2.7	10.4	52.7	34.1	26.4	14.8	31.9	26.9
Eroski	0	3.3	8.8	87.9	15.4	34.1	34.6	15.9
Mercadona	33.0	30.8	36.3	0	42.3	18.1	24.7	14.8

that no single supermarket yields utilities that are consistently higher or lower than the others. Although positions in the ranking of utilities seldom change daily, there is considerable longer-term fluctuation.

Table 4 reports the percentage of time that the price and the utility spend in each quartile (at the Barcelona location).

From the table's first four columns we observe that Auchan is the only supermarket that have occupied the first quartile for the entire study period; in other words, it was always among the lowest-priced stores. Mercadona was located in the two first quartiles for almost two thirds of the period. At the other extreme, Eroski was the most expensive supermarket 88% of the time and El Corte Inglés was among the least expensive only 2.7% of the time. Carrefour was usually located in the second quartile, and Condis was most often located in the third quartile. Looking at the distribution of utilities (last four columns of the table), we can see that there is less concentration in particular quartiles because the distribution of utilities is more spread out. These observed patterns support the notion that price dispersion is due to mixed strategies in combination with chain heterogeneity. In other words, each chain has its own price distribution from which to draw and—depending on the extent of firm heterogeneity—the respective supports of those distributions may overlap.

6.2 Search cost estimates for the basic basket

We use the basket prices to estimate the model's parameters.²⁵ We are mainly interested in estimating the proportion of consumers who search and calculating their search costs, that is, $\{q_k, c_k\}_{k=1}^n$. The estimation results, which are obtained using the ML procedure described previously, are presented in Table 5.

²⁵ One assumption of the theoretical model is that price observations could be uncorrelated. We check for serial correlation and calculate the autocorrelation function for the basket in each store–location pair. Not surprisingly, the autocorrelation between two consecutive days is fairly high; yet autocorrelations calculated with a 10-day lag between price observations are not significantly different from zero. So when estimating search costs, we use price observations separated by 10 days.

Table 5 Estimation results

	Barcelona	Madrid	Málaga	Vigo
q_1	0.838 (0.08)	0.717 (0.08)	0.786 (0.08)	0.588 (0.08)
q_2	0.119 (0.05)	0.199 (0.04)	0.151 (0.04)	0.281 (0.03)
⋮				
q_{10}	0.026 (0.03)	0.039 (0.05)	0.044 (0.04)	0.023 (0.06)
Price (mean)	66.79	66.38	66.79	66.24
Price (min.)	61.52	59.52	62.10	58.44
Price (max.)	69.32	69.32	69.32	69.32
$v_j - r_j$	9.72	7.52	6.64	5.78
LL	114.91	109.30	78.89	111.12
N	10	10	10	10
R^2	0.919	0.952	0.933	0.946
Observations	114	95	95	95

We first estimate utilities by running the chain–fixed-effects regression of prices separately for each location. The resulting R^2 values indicate that, in all locations, more than the 90% of the variation in prices is explained by chain dummies.

The first three rows of Table 5 report the estimated proportion of consumers searching one time (q_1) or two times (q_2). The estimated share of consumers who search once or twice is about 90% in all locations, but the percentage of consumers who search *all* local stores is never more than 4.4%. These results might indicate that there are no significant differences in search costs among the respective consumers from different locations, although the proportion of those who search only once is lower in Vigo and higher in Barcelona. Search costs may reflect the opportunity cost of time (which directly affects the cost of acquiring information) and/or other consumer characteristics (e.g., education, age).

Unfortunately we cannot analyze how search costs vary among consumers within each city as we have not information about consumer characteristics. For example, Dubois and Perrone (2015) find that, across income levels, there are no consistent differences in the proportion of individuals who engage in searching when shopping. On the other hand, De los Santos (2017) using data on consumers shopping for books online finds that consumers with higher-income individuals devoting less time to search.

Our results are in line with those reported in other papers. With regard to UK grocery items, Wildenbeest (2011) finds that most of the observed price variation is explained by supermarket heterogeneity and that the estimated amount of searching is low. He reports that 71% of consumers search only once, 91% search either once or twice, and only 8% of consumers compare all prices. Richards et al. (2016) show that 84% of consumers search only one store when shopping for products in many categories at the same time. Using data from France on food expenditures, Dubois and Perrone (2015) confirm that consumers observe only a few prices before making a purchase.

Table 6 Price of the occasional basket: alcoholic drinks

Supermarket	Price			
	Mean	SD	Max.	Min.
Auchan	66.74	1.59	70.28	62.59
Carrefour	73.40	1.43	75.69	70.90
Condis	72.93	1.75	76.25	70.56
ElCorteInglés	73.60	1.15	75.35	69.38
Eroski	73.09	1.76	76.28	69.26
Mercadona	73.46	1.81	76.24	70.44
Average	72.10	3.05	76.28	62.59

Estimated search costs are low. We estimate that, the search cost of a consumer who searches just once range from €0.45, in Málaga and €0.61 in Madrid. So for these consumers we can only infer that their search cost should have been at least of this amount in order to rationalize their behavior. Similarly, for consumers who search twice the search costs should have been at least €0.27 in Málaga and €0.37 in Madrid. Consumers who are comparing more than two prices should have been lower search costs. Such low search costs are similar to those found by Wildenbeest (2011), who calculates that the search cost of consumers who do *not* search should have been at least €0.27 for that behavior to be rational.

The estimated maximum price–cost margins ($v_j - r_j$) range from €5.78 in Vigo to €9.72 in Barcelona, which translates into a maximum margin below 10% in all locations—except in Barcelona. We remark that, in this last location, consumers search less than elsewhere, which could explain the higher margins in that city. Estimated margins in the United Kingdom range between 8 and 9%.

6.3 Price dispersion: occasional basket

Our results could be affected by the particular products selected for the shopping basket. All the products in our basic basket are among the most frequently purchased, so it is fair to suppose that consumers have more information about them. We therefore put together a substantially different basket containing only alcoholic beverages (wine, beer, and spirits). These products are assumed to be purchased less frequently, and most of them are not included in the basket basket.²⁶

The average cost of this basket is approximately €72, and its price dispersion is greater than that of the basic food basket. As Table 6 shows, the price ranges from €62.59 at Auchan to €76.28 at Eroski—a difference of almost 20%. Auchan remains

²⁶ Table 10 gives the average price (and price dispersion) of each product in this basket. The basket price is the sum of all individual prices weighted by subcategory (according to the Spanish Household Budget Survey, 2013). The share of expenditures on these three groups is 38.7% (beer), 43.9% (wine) and 17.4% (spirits). The specific items included in each category were chosen among those available in the six supermarkets at the four locations.

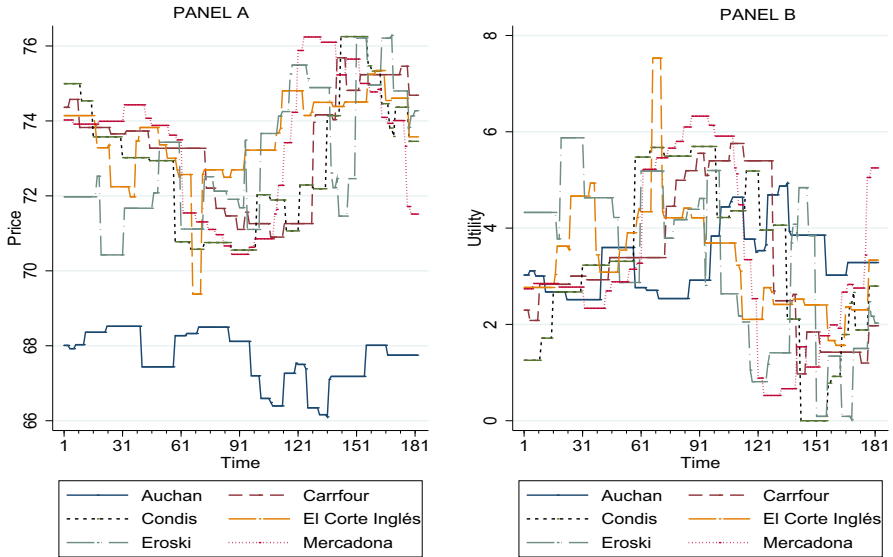


Fig. 3 Ranking over time: occasional basket (Barcelona)

the supermarket with the lowest prices (just as for the basic basket), but now Eroski (rather than El Corte Inglés; cf. Table 3) is, on average, the most expensive supermarket.

Figure 3 plots (once again, for Barcelona) the evolution of our occasional basket’s price and utility. The patterns are strongly similar to those observed for the basic basket (cf. Fig. 2). Although Auchan is again always the least expensive supermarket, there is variability in the price ranking of the other supermarkets. Panel B of the figure shows that utility rankings (calculated as the negative of the residuals) change more frequently than do prices and that no supermarket exhibits a consistent ranking.

6.4 Search cost estimates for the occasional basket

We first regress prices on store fixed effects to obtain the utilities. For most locations, these fixed effects explain more than 75% of observed price variability, as Table 7 shows.

The estimated portion of consumers who search only once ranges from 65% to 71%, although the differences are not statistically significant among cities. Search costs are higher than for the basic basket. We estimate that, the search cost of a consumer who searches just once range from €1, in Málaga and €1.1 in the rest of locations. So consumers search cost should have been larger than this amount in order to rationalize their behavior.

The maximum estimated price–cost margin ranges between 15 and 18%, values that are also higher than for the basic basket.

So in comparison with the basic basket, the occasional basket is characterized by greater price dispersion, higher search costs, and higher margins. These results accord with those in Sorensen (2000), who establishes that prices of repeatedly purchased

Table 7 Estimation results: occasional basket

	Barcelona	Madrid	Málaga	Vigo
q_1	0.647 (0.08)	0.702 (0.11)	0.710 (0.10)	0.689 (0.10)
q_2	0.264 (0.04)	0.248 (0.07)	0.224 (0.05)	0.250 (0.06)
⋮				
q_{10}	0.090 (0.05)	0.050 (0.04)	0.066 (0.05)	0.062 (0.05)
Price (mean)	72.29	71.82	72.11	72.03
Price (min.)	66.35	62.59	65.52	64.66
Price (max.)	76.25	76.24	76.24	76.28
$v_j - r_j$	10.87	12.99	11.58	12.16
LL	206.3	178.3	164.7	174.2
N	10	10	10	10
R -square.	0.657	0.776	0.739	0.746
Observations	114	95	95	95

prescriptions—for which search benefits are expected to be high—exhibit relatively low price–cost margins and relatively less price dispersion.

7 Conclusions

The aim of this empirical paper is to identify patterns of cross-sectional and temporal price dispersion, and to assess the role played by search costs and chain differentiation in that dispersion. We build a price data set of grocery and household products often included on Spanish consumers' shopping lists and sold at the main Spanish supermarket chains.

Quantifying search costs is a relevant question from the perspective of competition policy. As Watson (2003) shows, traditional policies that fail to incorporate search costs will likely be less effective than other policies at enhancing competition. In fact, consumer information about prices is a necessary condition for markets to be truly competitive. If consumers never engaged in price comparison, then the monopoly price would prevail in equilibrium (Diamond 1971); however, price dispersion will be an equilibrium in markets where at least some consumers search more than once (Burdett and Judd 1983). Regulations that aim to promote competition must therefore account for the distortions due to informational restrictions. For instance, Stahl (1989) shows that—in the presence of search costs—firm entry does not necessarily improve consumer welfare. Similarly, Lach and Moraga-González (2017) find that consumer surplus always (although weakly) decreases with increased competition. For those agencies tasked with devising competition policy, another consideration should be retailer practices that aim to confuse or mislead consumers (Ellison and Ellison 2009).

Analyzing consumer search in online markets could also yield insights into traditional, brick-and-mortar retail behavior. Consumer access to prices through supermarket websites or price comparison sites (e.g., Soysuper.com) has reduced search

costs for both online and offline retail consumers. Customers can easily check prices before they shop and also while they shop.

We find that some chains have persistently lower prices than others, even as prices change frequently. This empirical evidence suggests that both search frictions and chain differentiation help explain price dispersion in Spanish grocery markets. We estimate the distribution of consumer search costs for different baskets of goods sold in different geographical markets. For that purpose we employ the model proposed by Burdett and Judd (1983) and modified—to account for vertical product differentiation—by Wildenbeest (2011). We find (for Barcelona) that about 90% of the observed variation in prices is due to chain fixed effects and that the estimated share of consumers searching only once is 84% and such behavior is economically rational only when search costs amounted to at least €0.57. Our regression results also indicate that the more frequently purchased products tend to have lower price dispersion and lower price–cost margins.

Finally, our results are in line with findings previously reported in the literature on the retail food market in other countries—for instance, France and the United Kingdom. Moreover, are consistent with the Ministry of the Environment and Rural and Marine Affairs (MARM 2011) survey where 84% of interviewed Spanish consumers consider themselves to be fairly loyal to a supermarket and 48.4% of them always buy without comparing prices.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix

See Tables 8, 9 and 10

Table 8 Number of products (observations) in each location and supermarket

Location	Auchan	Carrefour	Condis	El Corte Inglés	Eroski	Mercadona	Total
Barcelona	223 (40,473)	234 (42,588)	200 (36,352)	235 (42,770)	237 (43,134)	173 (31,486)	1302 (236,803)
Madrid	223 (40,477)	233 (42,406)	–	215 (39,130)	236 (42,952)	182 (33,124)	1089 (198,089)
Málaga	221 (40,001)	233 (42,406)	–	235 (42,770)	237 (43,134)	181 (32,942)	1107 (201,253)
Vigo	221 (40,002)	231 (42,042)	–	236 (42,952)	235 (42,723)	177 (32,210)	1100 (199,929)
Total	888 (160,953)	931 (169,442)	200 (36,352)	921 (167,622)	945 (171,943)	713 (129,762)	9494 (836,074)

Table 9 Summary statistics for items in the basic basket

	Mean price (SD)	Minimum price	Maximum price	Coef. of Var. ($\times 100$)
Beverages				
Aquarius lemon 1.5 l	1.41 (0.10)	0.97	1.49	6.74
Coca Cola 2 l	1.36 (0.02)	1.01	1.38	1.57
Mineral water Aquarel Nestle 5 l	1.03 (0.07)	0.84	1.21	6.71
Beer Heineken pack 6 \times 25 cc.	3.23 (0.33)	1.97	3.65	10.29
Wine Rioja El Coto 75 cl.	5.17 (0.15)	3.82	5.38	2.96
Breakfast and cereals				
Cereal Bar Chocho Crispies 6 \times 20 g	2.12 (0.13)	1.61	2.29	6.24
Corn Flakes Kelloggs 500 g	1.98 (0.09)	1.37	2.25	4.78
Special K Kelloggs 500 g	2.82 (0.22)	1.87	3.00	7.86
All bran Kelloggs cereal bar 6 \times 40 g	2.36 (0.12)	1.61	2.50	5.00
Breakfast Biscuits Flora 450 g	1.39 (0.11)	1.16	1.59	8.05
Breakfast Biscuits María Fontaneda	1.98 (0.12)	1.58	2.1	6.17
Dairy products				
Low-fat milk Asturiana 4 \times 1.5 l	2.74 (0.06)	2.48	2.78	2.10
Yogurt Activia Natural	1.78 (0.06)	1.18	1.80	3.57
Yogurt Danone Pack 8 \times 125 g	1.85 0.07	1.49	1.89	3.93
Yogurt low-fat Vitalinea Danone 4 \times 125 ml	0.99 (0.03)	0.90	1.26	2.58
Condensed milk La Lechera	2.19 0.12	1.87	2.65	5.69
Pantry				
Coffee capsules 10u.	3.25 (0.11)	2.44	2.29	3.44
Mermelade Hero 350 g	1.67 (0.13)	1.09	1.95	7.69
Cocoa cream Nocilla 400 g	2.69 (0.13)	2.29	2.89	4.90
Rice Brillante 1k.	1.59 (0.06)	1.28	1.67	3.77
Pasta macaroni (Gallo)	0.76 (0.05)	0.62	0.80	6.14
Lentils (Litoral)	1.51 (0.15)	1.14	2.56	9.97
Chicken Broth Gallina Blanca 1 l	1.90 (0.11)	1.39	1.95	6.00
Corn Flour (Maizena) 350 g	1.41 (0.14)	1.02	1.69	9.90
Chocolate with almond (Nestlé) 150 g	1.26 (0.12)	0.96	1.49	9.62
Personal care and household				
Deodorant Roll on (Dove)	2.14 (0.39)	1.62	2.95	18.32
Champoo Total Repair 5 Elvive Loreal	2.87 (0.18)	1.96	3.05	6.25
Champoo Pro-v Pantene	2.96 (0.10)	2.09	3.19	3.21
Hair Mask Pantene	4.66 0.21	3.52	4.9	0.04
Cillit bang 750 ml	3.49 0.25	1.99	3.89	0.07
Entire Basket	79.08 (2.25)	70.95	82.08	2.85

All prices are given in euros

Table 10 Summary statistics for items in the occasional basket

	Mean price (SD)	Minimum price	Maximum price	Coefficient of variation ($\times 100$)
Beer				
San Miguel lemon Beer 6 \times 25 cl.	2.20 (0.17)	1.42	2.49	7.67
Voll-damm Beer 6 \times 25 cl.	3.59 (0.26)	2.70	3.79	7.32
Heineken Beer 6 \times 25 cl.	3.23 (0.33)	1.97	3.65	40.29
Wine				
Mateus Rose Wine 75 cl.	3.21 (0.17)	1.42	4.15	14.41
El coto Rioja Wine 75 cl.	5.17 (0.26)	2.70	5.38	2.95
Valdepeñas Wine 75 cl.	3.17 (0.33)	1.97	3.35	6.06
Freixenet Cava Carta Nevada 75 cl.	4.96 (0.46)	3.56	6.59	8.24
Martini Rosso Vermouth 1 l.	8.21 (0.15)	7.25	5.45	3.49
Cava Anna de Codorniu 75 cl.	8.15 (0.19)	5.99	8.49	3.75
Spirits				
Vodka Absolut 70 cl.	11.75 (0.27)	10.09	12.45	2.36
Ron Brugal 70 cl.	12.74 (0.31)	11.79	12.95	2.46
Whisky J.B. 70 cl.	11.62 (0.37)	10.49	13.02	3.20
Gin Beefeater London Dry 70 cl.	12.57 (0.37)	10.51	13.79	2.97
Whole Basket	72.10 (3.05)	76.28	62.59	4.22

All prices are given in euros

References

- Bachis E, Piga CA (2011) Low-cost airlines and online price dispersion. *Int J Ind Organ* 29(6):655–667
- Barron JM, Taylor BA, Umbeck JR (2004) Number of sellers, average prices and price dispersion. *Int J Ind Organ* 22(8–9):1041–1066
- Baye MR, Morgan J (2001) Information gatekeepers on the Internet and the competitiveness of homogeneous product markets. *Am Econ Rev* 91(3):454–474
- Baye MR, Morgan J, Scholten P (2004a) Temporal price dispersion: evidence from an online consumer electronics market. *J Interact Mark* 18(4):101–115
- Baye MR, Morgan J, Scholten P (2004b) Price dispersion in the small and in the large: evidence from an Internet price comparison site. *J Ind Econ* 52(4):463–496
- Baye MR, Morgan J, Scholten P (2006) Information, search and price dispersion. In: Hendershott T (ed) *Handbook of economics and information systems*, vol 1. Elsevier Press, Amsterdam, pp 323–371
- Baye MR, De los Santos B, Wildenbeest MR (2013) The evolution of product search. *J Law Econ Policy* 9:201–221
- Berck P, Brown J, Perloff JM, Villas-Boas SB (2008) Sales: test of theories on causality and timing. *Int J Ind Organ* 26(6):1257–1273
- Brown JR, Goolsbee A (2002) Does the Internet make markets more competitive? Evidence from the life insurance industry. *J Polit Econ* 110(3):481–507
- Brynjolfsson E, Smith MD (2000) Frictionless commerce? A comparison of Internet and conventional retailers. *Manag Sci* 46(4):563–585
- Brynjolfsson E, Dick AA, Smith MD (2010) A nearly perfect market? Differentiation versus price in consumer choice. *Quant Mark Econ* 8:1–33
- Burdett K, Judd K (1983) Equilibrium price dispersion. *Econometrica* 51(4):955–969
- Carlson JA, McAfee RP (1983) Discrete equilibrium price dispersion. *J Polit Econ* 91(3):480–493

- Cavallo A (2017) Are online and offline prices similar? Evidence from multi-channel retailers. *Am Econ Rev* 107(1):283–303
- Cavallo A (2018) Scraped data and sticky prices. *Rev Econ Stat*. https://doi.org/10.1162/REST_a_00652
- Chandra A, Tappata M (2011) Consumer search and dynamic price dispersion: an application to gasoline markets. *RAND J Econ* 42(4):681–704
- Clay K, Krishnan R, Wolff E (2001) Prices and price dispersion on the web: evidence from online book industry. *J Ind Econ* 49(3):521–539
- Comisión Nacional de la Competencia (CNC) (2011) Report on the relations between manufacturers and retailers in the food sector
- Dana JD (1994) Learning in an equilibrium search model. *Int Econ Rev* 35(3):745–771
- De los Santos B (2017) Consumers search on Internet (Manuscript). http://babur.people.clemson.edu/research/Consumer_Search_on_the_Internet.pdf. Accessed 23 Jan 2018
- De los Santos B, Hortaçsu A, Wildenbeest MR (2012) Testing models of consumer search using data on web browsing and purchasing behavior. *Am Econ Rev* 102(6):2955–2980
- Delgado J, Waterson M (2003) Tyre price dispersion across retail outlets in the UK. *J Ind Econ* 51(4):491–509
- Diamond PA (1971) A model of price adjustment. *J Econ Theory* 3(2):156–158
- Dubois P, Perrone H (2015) Price dispersion and search costs: the roles of imperfect information and product differentiation (Manuscript)
- Edelman B (2012) Using Internet data for economic research. *J Econ Perspect* 26(2):189–206
- Ellison G, Ellison SF (2009) Search, obfuscation and price elasticities on the Internet. *Econometrica* 77(2):427–452
- González X (2017) Chain heterogeneity and price-setting behavior: evidence from e-grocery retailers. *Electron Commer Res Appl* 26:62–72
- Hong H, Shum M (2006) Using price distributions to estimate search costs. *RAND J Econ* 37(2):257–275
- Hopkins E (2008) Price dispersion. In: Durlauf SN, Blume LE (eds) *The new palgrave dictionary of economics*. Palgrave Macmillan, London
- Hortaçsu A, Syverson C (2004) Product differentiation, search costs and competition in the mutual fund industry: a case study of S&P 500 index funds. *Quart J Econ* 119(2):403–456
- Janssen M, Moraga-González JL (2004) Strategic pricing, consumer search and the number of firms. *Rev Econ Stud* 71(4):1089–1118
- Janssen M, Moraga-González JL, Wildenbeest MR (2005) Truly costly sequential search and oligopolistic pricing. *Int J Ind Organ* 23(5–6):451–466
- Janssen M, Moraga-González JL, Wildenbeest MR, Brousseau E, Curien N (2007) Consumer search and pricing behavior in Internet markets. In: Brousseau E, Curien N (eds) *Internet and digital economics*. Cambridge University Press, Cambridge, pp 460–483
- Lach S (2002) Existence and persistence of price dispersion: an empirical analysis. *Rev Econ Stat* 84(3):433–444
- Lach S, Moraga-González JL (2017) Asymmetric price effects of competition. *J Ind Econ* 65(4):767–803
- MAPAMA (2014) Informe sobre el consumo alimentario en España, 2014. Ministerio de Agricultura, Pesca, Alimentación y Medio Ambiente. http://www.mapama.gob.es/es/alimentacion/temas/consumo-y-comercializacion-y-distribucion-alimentaria/dic14_tcm7-382143.pdf. Accessed 29 Sept 2017
- MARM (2011) Estudio de mercado. Observatorio del consumo y la distribución alimentaria, Ministerio de Medio Ambiente y Medio Rural y Marino
- Moraga-González JL, Wildenbeest MR (2008) Maximum likelihood estimation of search costs. *Eur Econ Rev* 52(5):820–848
- Moraga-Gonzalez JL, Sandor Z, Wildenbeest M (2013) Semi-nonparametric estimation of consumer search costs. *J Appl Econ* 28:1205–1223
- Moraga-González JL, Sandor Z, Wildenbeest M (2017) Non-sequential search equilibrium with search cost heterogeneity. *Int J Ind Organ* 50:392–414
- Orlov E (2011) How does the Internet influence price dispersion? Evidence from the airline industry. *J Ind Econ* 59(1):21–37
- Richards T, Hamilton SF, Allender W (2016) Search and price dispersion in online grocery markets. *Int J Ind Organ* 47:255–281
- Rosenthal RW (1980) A model in which an increase in the number of sellers leads to a higher price. *Econometrica* 48(6):1575–1579

- Seiler S (2013) The impact of search costs on consumer behaviour: a dynamic approach. *Quant Mark Econ* 11:155–203
- Sorensen AT (2000) Equilibrium price dispersion in retail markets for prescription drugs. *J Polit Econ* 108(4):833–850
- Stahl DO (1989) Oligopolistic pricing with sequential consumer search. *Am Econ Rev* 79(4):700–712
- Stigler G (1961) The economics of information. *J Polit Econ* 69(3):213–225
- Varian HR (1980) A model of sales. *Am Econ Rev* 70(4):651–659
- Waterson M (2003) The role of consumers in competition and competition policy. *Int J Ind Organ* 21(2):129–150
- Wildenbeest MR (2011) An empirical model of search with vertically differentiated products. *RAND J Econ* 42(4):729–757
- Zhao Y (2006) Price dispersion in the grocery market. *J Bus* 79(3):1175–1192

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.